HHS Public Access

Author manuscript

Cell Rep. Author manuscript; available in PMC 2017 January 26.

Published in final edited form as:

Cell Rep. 2016 November 01; 17(6): 1607–1620. doi:10.1016/j.celrep.2016.10.011.

Widespread chromatin accessibility at repetitive elements links stem cells with human cancer

Nicholas C. Gomez[1,2,3], **Austin Hepperla**[1,2,3], **Raluca Dumitru**[2,4,5], **Jeremy M. Simon**[1,2,6], **Fang Fang**[1,2], and **Ian J. Davis**[1,2,7,*,†]

[1]Lineberger Comprehensive Cancer Center, University of North Carolina, Chapel Hill, NC, 27599

[2]Department of Genetics, University of North Carolina, Chapel Hill NC, 27599

[3]Curriculum in Genetics and Molecular Biology, University of North Carolina, Chapel Hill NC

[4]Neuroscience Center, University of North Carolina, Chapel Hill, NC, 27599

[5]Human Pluripotent Stem Cell Core Facility, University of North Carolina, Chapel Hill NC, 27599

[6]Carolina institute for Developmental Disabilities, University of North Carolina, Chapel Hill NC, 27599

[7]Department of Pediatrics, University of North Carolina, Chapel Hill NC, 27599

Summary

Chromatin regulation is critical for differentiation and disease. However, features linking the chromatin environment of stem cells with disease remain largely unknown. We explored chromatin accessibility in embryonic and multipotent stem cells and unexpectedly identified widespread chromatin accessibility at repetitive elements. Integrating genomic and biochemical approaches, we demonstrate that these sites of increased accessibility are associated with well-positioned nucleosomes marked by distinct histone modifications. Differentiation is accompanied by chromatin remodeling at repetitive elements associated with altered expression of genes in relevant developmental pathways. Remarkably, we found that the chromatin environment of Ewing sarcoma, a mesenchymally derived tumor, is shared with primary mesenchymal stem cells (MSC). Accessibility at repetitive elements in MSC offers a permissive environment that is exploited by the critical oncogene responsible for this cancer. Our data demonstrate that stem cells harbor a unique chromatin landscape characterized by accessibility at repetitive elements, a feature associated with differentiation and oncogenesis.

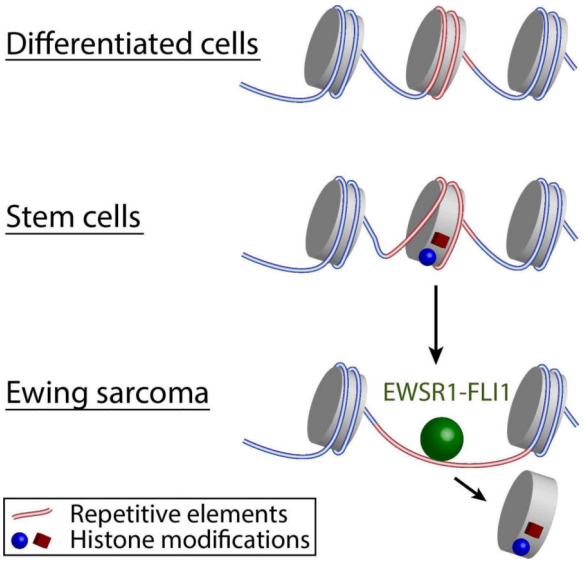
[*]Correspondence to: UNC Lineberger, 450 West Drive, Chapel Hill, 27599, 919-966-5360, ian_davis@med.unc.edu.
[†]Lead contact

Author Contributions

N.G.: conception and design, collection and assembly of data, data analysis and interpretation, and manuscript preparation; R.D. collection of data; F.F., A.H. and J.S: bioinformatics support and analyses; I.D.: conception and design, data interpretation, financial support, manuscript preparation, and final approval of manuscript.

Accession Number: GSE75172

Graphical Abstract



Keywords

Stem cell; chromatin; repetitive elements; histone modification; chromatin accessibility; differentiation; Ewing sarcoma

Introduction

Early mammalian development necessitates precisely regulated transcriptomic and chromatin changes as cells commit to their terminal fates (Dixon et al., 2015; Kurimoto et al., 2015; Paige et al., 2012; Wamstad et al., 2012). A comprehensive understanding of chromatin remodeling during differentiation may reveal biological pathways that regulate this process and could suggest therapeutic opportunities relevant to cancer-directed and regenerative medicine. Human embryonic stem cells (hESCs), derived from the inner cell mass of human blastocysts, can propagate *in vitro* and are able to undergo multi-lineage differentiation (Thomson et al., 1998). Previous studies have explored chromatin dynamics during stem cell differentiation by comparing hESCs to differentiated cells. hESCs are characterized by elevated levels of activation-associated histone post-translational modifications, histone bivalency at developmentally regulated genes, and increased expression of variant histones (Bernstein et al., 2006; Kafer et al., 2010; Mikkelsen et al., 2007; Wen et al., 2009). Though insightful, histone modification changes represent one of multiple strategies that ultimately regulate the chromatin landscape.

Ewing sarcoma is a highly malignant tumor of the bone and soft tissue with a peak incidence during adolescence. This tumor is virtually always characterized by a recurrent chromosomal rearrangement that brings together the amino terminus of EWSR1 with the carboxyl DNA binding domain of the ETS family transcription factor FLI1. We and others have shown that the chimeric oncoprotein is selectively targeted away from canonical ETS sites to coopt microsatellite repeats that contain the core recognition element sequence

(Gangwal et al., 2008; Patel et al., 2012). At these sites EWSR1-FLI1 is necessary to maintain a fully accessible chromatin landscape marked by enhancer associated histone modifications (Patel et al., 2012; Riggi et al., 2014). Many of the genes implicated in tumor development and regulated by EWSR1-FLI1 are located proximally to these microsatellite repeats (Grunewald et al., 2015; Kinsey et al., 2006; Luo et al., 2009). Despite its chromatin remodeling activity, EWSR1-FLI1 only demonstrates cancer-like targeting in Ewing sarcoma cells. What mediates the selective targeting of EWSR1-FLI1 and what this indicates about the cell-of-origin remain unknown.

In an effort to comprehensively explore features of chromatin organization that accompany early mesenchymal differentiation and a potential association with Ewing sarcoma, we utilized FAIRE-seq, an unbiased biochemical assay that enriches for localized regions of nucleosome-depleted ("open") chromatin (Giresi et al., 2007; Simon et al., 2012). Regions identified by FAIRE-seq include a broad range of regulatory classes. We applied this technique to compare the chromatin landscape of hESC, primary and *in vitro* differentiated mesenchymal stem cells and mature cell lines. We identified increased chromatin accessibility at specific classes of repetitive elements in stem cells. These regions harbored distinct histone modifications and underwent chromatin remodeling during differentiation. A subset of repetitive elements exhibiting enhanced chromatin accessibility in stem cells offered a permissive environment that could be exploited by EWSR1-FLI1 in Ewing sarcoma lending support of a stem cell origin for this cancer and offering a mechanistic explanation for its selective targeting.

RESULTS

FAIRE-selected chromatin from human embryonic stem cells is dominated by repetitive elements

To explore chromatin organization in human embryonic stem cells, we performed FAIRE-seq on undifferentiated H1-ESC (WA01), H7-ESC (WA07), and H9-ESC (WA09) cells and aligned sequencing reads to the human genome, as previously described (Langmead et al., 2009)(Simon et al., 2014). As expected, FAIRE signal was enriched at transcriptional start sites (TSS) and CTCF binding sites in all hESC (Figure S1A) (Simon et al., 2014). We also observed signal enrichment at OCT4 and NANOG binding sites, factors critical for the maintenance of pluripotency (Figure S1A) (Boyer et al., 2005; Loh et al., 2006).

We then identified genomic regions that were unique to stem cells. We compared z-score-transformed FAIRE signal in 500 bp windows to publicly available data from three differentiated cell types, each representing distinct developmental lineages (HUVEC, K562, and NHEK) (Consortium et al., 2012). Of the regions that passed a minimum signal filter, 12,026 sites demonstrated a significant difference between hESC and the three differentiated cell types ($p <= 0.01$, t-test). Hierarchical clustering resolved these regions into two major groups (Figure 1A). Cluster 1 (C1) consisted of regions with increased FAIRE signal in hESCs. Cluster 2 (C2) contained regions with higher signal in the differentiated cell lines (Figure 1A and B). The two clusters demonstrated significant differences in location. C1 regions were primarily distal, with a median distance of 39.5 Kb to the nearest TSS. Whereas C2 regions were primarily proximal, with a median distance of 11.4 Kb (Figure

1A). We then annotated the genomic intervals with classifications previously generated by segmentation analyses in H1-ESC, HUVEC, K562, and NHEK (ChromHMM) (Consortium et al., 2012; Ernst and Kellis, 2010, 2012). C1 was significantly enriched for transcription and heterochromatic/repetitive states ($p < 0.001$, Figure 1A and Figure S1B). In contrast, C2 was enriched for states such as active and poised promoters, as well as insulators. ($p < 0.001$, Figures 1A and S1B). Interestingly, despite the striking difference in FAIRE signal between cell types, regions in these clusters were similarly classified. Taken together, these data revealed widespread accessible chromatin in stem cells at genomic regions classified as heterochromatic.

To further characterize FAIRE selected chromatin, we identified 610,887 regions with significant signal enrichment (peaks) in H1-ESC, 243,467 in H7-ESC, and 384,162 in H9-ESC (MACS2) (Zhang et al., 2008). Applying a false discovery rate threshold, we selected the top ~150,000 peaks for further analysis. The filtered regions were then intersected with repetitive elements defined by RepeatMasker (Smit et al., 2015), requiring that the site of greatest FAIRE signal was within one bp of a repetitive element. Strikingly, we found that over 82.9%, 94.6% and 94.0% of peak summits identified in H1-ESC, H9-ESC, and H7-ESC, respectively, intersected a repetitive element. The degree of overlap for each hESC was significantly greater than HUVEC, NHEK, K562, and a randomly permuted peak set (Figure 1C). Varying the stringency used to select peaks had no effect on fractional overlap (Figure S1C).

Simple repeats and SINEs are enriched in FAIRE selected chromatin

We then assessed whether the enrichment of repetitive elements was restricted to specific classes. Simple repeats and short interspersed nucleotide elements (SINEs) were selectively enriched among FAIRE peaks in hESC, relative to their genomic prevalence. This pattern was not observed in the three differentiated cell types (Figure 1D).

In further support of repetitive element enrichment, we also found that a large fraction of sequencing reads from each hESC line was discarded during alignment due to redundant genomic mapping (Table S1). 83% of unaligned sequences from H1-ESC were repetitive in nature, enriched for SINEs, simple repeats, and LINEs (Table S2). In contrast, similar analysis of HUVEC FAIRE identified only 51% of discarded reads as repetitive sequences, a fraction consistent with the abundance of these elements genome-wide.

We then assayed FAIRE signal differences in each repeat class. After normalizing for repeat length and sequencing depth, signal in hESC at simple repeats and SINEs greatly exceeded that of differentiated control cells (Figure 1E). In contrast, signal differences at LINEs were minimal and DNA transposons demonstrated an inverse relationship. Taken together, read, signal and peak-based detection approaches consistently identify the selective enrichment of simple repeats and SINEs by FAIRE in hESCs.

FAIRE-enriched repeats are shared across hESCs and exhibit distinguishing features

Simple repeats and SINEs consist of thousands to millions of individual regions. Only a small fraction was identified by FAIRE (< 5% of each class, Table S3). Since it might be expected that repetitive regions of the genome pose challenges to accurate sequence

alignment, we explored the ability of sequencing reads to align to repetitive elements. Using on a 50 bp Kmer, we found that 78% of all nucleotide positions within repetitive elements are mappable using default alignment criteria. Setting a more stringent threshold by permitting only unique reads to map, 72% of all base pairs were deemed unique. Further, using the ENCODE mappability track, we found that the majority (51%) of FAIRE positive SINEs contain enough sequence diversity to map 50 bp reads in over 50% of the repeat length and contain bps near the 5' and 3' ends that have an average mappability score > 0.5 (Figure S1D, E). Taken together, these data demonstrate that despite classification as repetitive, these regions contain sufficient sequence diversity to enable accurate mapping of sequence tags. However, it remains possible that variation in mappability across individual repetitive elements may lead to over or underestimation of FAIRE signal at specific regions.

We then asked if simple repeats and SINEs were consistently identified across hESC lines. We observed a significant overlap of simple repeats and SINEs with signal in the top quartile between hESC line (simple repeat: 48%, 86949 of 179379; SINEs: 59%, 423,819 / 723415; $p < 0.001$ by permutation) (Figures 2A, D, and S2A, S2B). Consistent with a central role of the repetitive segment in mediating chromatin state, we found that for both simple repeats and SINEs FAIRE signal was centered at the repetitive element, rather than extending from flanking regions, and was concordant between the stem cells (Figures 2B, E and S2C, S2D).

To explore other factors that may influence chromatin status, we asked whether length and G/C content distinguish those repeats that are FAIRE-enriched. For simple repeats, the lengths of FAIRE-enriched and FAIRE-negative sites varied little. However, enriched sites demonstrated a significant skew towards higher G/C content (Figure 2C). The opposite pattern was observed for SINEs. FAIRE-enriched SINEs were significantly longer than others whereas G/C content differed only slightly (Figure 2F). Overall, FAIRE identifies repetitive elements that are common to multiple hESCs and demonstrate shared chromatin patterns and distinct class-specific DNA features.

FAIRE and DNase differ at repetitive regions

Given the abundance of FAIRE-enriched repeats, we were surprised that chromatin accessibility at these sites had not been previously observed. As a complementary approach, we analyzed DNase hypersensitivity data (DNase). In contrast to FAIRE, DNase depends on enzymatic digestion to interrogate chromatin accessibility (Boyle et al., 2008; Buenrostro et al., 2013; Crawford et al., 2006). Leveraging publicly available data, we compared FAIRE and DNase signal at simple repeats and SINEs at FAIRE peaks (from Figure 1C). Surprisingly, these regions lacked DNase signal (Figure 3A). Conversely, the few repeats that demonstrated DNase signal lacked FAIRE enrichment (1099 and 2033 simple repeats and SINEs, respectively). As a control, we examined FAIRE and DNase at transcription start sites (TSS) and CTCF sites. FAIRE and DNase positively correlated at these regions, consistent with published results and confirming the validity of the assays (Figures 3B and S3A). To determine if the variation observed between FAIRE and DNase was due to differences in alignment of the shorter DNase reads, we truncated 50 bp FAIRE-seq reads to the 20-bp sequence used for DNase-seq and realigned them to genome. We again noted

enrichment of FAIRE signal at repetitive elements indicating that read length was not a factor (Figure S3B).

Because of the discrepancy between FAIRE and DNase at repetitive regions, we then explored nucleosome positioning using published MNase-seq data (West et al., 2014). By cleaving DNA in the linker region between two nucleosomes, MNase-seq offers insight into the location of nucleosomes. DNase-positive regions, including those in repeats, TSS, and CTCF binding sites, demonstrated decreased MNase signal, consistent with nucleosome depletion (Figure 3A). However, FAIRE-enriched SINEs and shorter simple repeats exhibited the presence of one to two well-positioned nucleosomes. Of note, phased nucleosomes flanked both classes of repeats, similar to patterns observed at other regulatory elements (Figure 3B) (Fu et al., 2008; Orvis et al., 2014).

To further characterize the relationship between FAIRE and nucleosome positioning at repetitive regions, we examined MNase signal at all simple repeats grouped by the magnitude of FAIRE enrichment. MNase signal was greatest at regions with highest FAIRE enrichment. Further, regions with the greatest FAIRE signal demonstrated the presence of a single centered nucleosome (Figure 3C). For all regions, we observed symmetrical nucleosome phasing extending beyond the repetitive region. Overall, these data indicate that, in contrast to the recognized association of FAIRE with nucleosome depletion, in the context of these regions, FAIRE identifies a chromatin organizational feature characterized by the presence of nucleosomes.

Distinct histone post-translational modifications demarcate repetitive elements

We then asked whether specific histone modifications distinguish the nucleosomes at accessible repetitive elements. We compared H1-ESC ChIP-seq data for a range of histone modifications at FAIRE-enriched and FAIRE-negative sites (Bernstein et al., 2010). We found that FAIRE-enriched simple repeats were marked by specific acetylated histones (Figure 4A). Associated modifications differed from those at FAIRE-enriched SINEs as well as TSS and CTCF sites (Figure 4A). H3K56ac and H2AK5ac were most associated with simple repeats. Signals for these modifications were centered over the repeat and demonstrated a magnitude similar or greater than that found at TSS and CTCF sites. (Figures 4B and S4A). H4K8ac and H2A.Z were most associated with SINEs and show subtle but center-weighted enrichment (Figures 4C and S4B). Overall, these data indicate that FAIRE-enriched simple repeats and SINEs are characterized by distinctly marked nucleosomes.

As an alternative approach to explore chromatin accessibility, we performed salt fractionation of MNase treated nuclei. Salt fractionation separates chromatin based on physical properties (Sanders, 1978). Low salt-soluble regions are enriched for active and highly accessible chromatin, whereas high salt solubilizes the bulk chromatin fraction (Henikoff et al., 2009). Salt fractionation both allows for direct comparisons of nucleosome composition in active chromatin as well as the positioning of individual nucleosomes by high-throughput sequencing. Nucleosomes were extracted from nuclei of H1-ESC and a differentiated control (Human Kidney Cells, HKC) using increasing concentrations of salt. The low salt fraction from both cell types consisted predominantly of mono-nucleosomes

whereas the high salt fraction consisted of mostly di-nucleosomes (Figure S5), consistent with published results (Teves and Henikoff, 2012). Histone post-translational modifications associated with each fraction were assayed by immunoblot. As predicted by our informatic analyses, H2AK5ac was significantly enriched in low salt fractions of nucleosomes from stem cells when compared to the differentiated control cells (p-value < 0.05, Figure 4D). This enrichment did not extend to the high salt or insoluble chromatin.

To identify nucleosome positioning at repetitive elements in highly accessible chromatin, we sequenced the DNA in both low and high salt soluble fractions and plotted the signal at simple repeats (Figure 4E). We again identified nucleosome phasing flanking the repeats in both H1-ESC and the differentiated control cells. However, compared with the differentiated cell control, H1-ESC demonstrated an increase in MNase signal at the center of the repeat exclusively in low salt extracted chromatin, indicative of a highly extractable nucleosome. Taken together with the immunoblot and ChIP-seq analysis, these data indicate that specific acetylation is associated with nucleosomal destabilization but not displacement at repetitive elements.

Repetitive regions undergo chromatin remodeling during differentiation

The difference in FAIRE enrichment at repetitive elements in stem and differentiated cells led us to test whether these elements undergo remodeling during differentiation. H1-ESC embryonic stem cells were differentiated in culture towards a mesenchymal lineage (H1-MSC). Differentiation of hESC to MSC was validated using several approaches. Morphologically, H1-MSCs acquired a fibroblastic appearance in contrast to the spherical colonies of H1-ESC (Figure S6A). The multipotency of H1-MSC was demonstrated by further differentiation into osteoblast and adipocyte lineages (Figure S6A). Finally, flow cytometry of H1-MSC identified a robust increase in CD90, CD73, CD105, and CD44, cell surface markers also detected on primary bone marrow-derived MSC (BM-MSC) (Figure S6B). H1-ESC were negative for CD73 and CD105.

FAIRE-seq was then performed on primary BM-MSC and H1-MSC. We identified ~15,000 SINEs and ~4,500 simple repeats with significantly different FAIRE signal between hESC and BM-MSC. Unsupervised hierarchical clustering of these regions, together with signal from H1-MSC as well as differentiated control cells, revealed two main clusters (Figure 5A and E). For SINEs, stem cells clustered together closely distinct from the differentiated cells. Notably, the differentiated H1-MSC exhibited greater similarity to primary BM-MSC than to undifferentiated H1-ESC. Overall, virtually all sites that exhibit signal variation demonstrated a progressive decrease in FAIRE enrichment accompanying differentiation. SINEs with differential FAIRE signal were then associated with the most proximal gene. Of those genes that were also differentially expressed, 95% demonstrate greater expression (> 2-fold) in hESC relative to BM-MSC, significant when compared to a permutation ($p<0.001$) (Figure 5B). The overall skew to greater message abundance in hESC is consistent with higher global transcription levels in these cells (Efroni et al., 2008). Genes with elevated expression in hESC (Category 1) were linked to curated ontologies related to ESC specific expression whereas those with elevated expression in MSC (Category 2) were

linked with terms implicating mesenchymal development such as wound repair and adipogenesis (Figures 5E and Table S4).

Clustering cell lineages based on FAIRE signal at simple repeats demonstrated a distinct pattern from that observed based on signal at SINEs. hESC clustered together, clearly separated from the differentiated cells. MSC, including bone marrow and H1-derived, clustered closely together but grouped with the differentiated cells. FAIRE signal at simple repeats revealed two patterns. One pattern was similar to that seen for the SINEs with progressively decreasing signal associated with differentiation. The other pattern revealed greater FAIRE signal in MSC lineages compared with either hESC or the differentiated cells (Figure 5E). We again associated these regions with differentially regulated genes. Regions with higher FAIRE signal in hESC or MSC were significantly associated with genes more highly expressed in hESC or MSC, respectively when compared to a permutation ($p = 0.02$ and $p < 0.001$ respectively) (Figure 5F). Similar to SINEs, hESC-associated genes were again enriched for gene ontologies related ESCs, whereas MSC-associated genes were enriched for pathways linked to mesenchymal development (Figures 5E and Table S4).

Taken together, these results suggest that repetitive elements undergo chromatin remodeling during differentiation. Repetitive regions with variable accessibility are associated with changes in lineage-specific gene expression and developmental pathways. However, the pattern of remodeling differs between the two classes of repetitive elements. SINEs primarily become inaccessible during differentiation, whereas a subset of simple repeats become accessible during lineage specification.

Oncogenic transcription coopts stem cell chromatin

Many sarcomas are thought to originate from stem cells of mesenchymal origin (Rodriguez et al., 2012). To explore this link, we compared the chromatin environment in stem cells with that in Ewing Sarcoma, the second most common bone malignancy in children and young adults. Ewing sarcoma is characterized by a chromosomal rearrangement that creates a chimeric transcription factor. We and others have previously shown that the resulting oncoprotein, EWSR1-FLI1, targets a subset of simple repeats distinct from the parental protein FLI1 (Gangwal et al., 2008). The binding of this transcription factor activates an oncogenic transcriptional profile critical for maintaining tumorigenicity (Gangwal et al., 2008). However, this targeting is cell-type specific (Patel et al., 2012). This observation led us to hypothesize that a permissive chromatin environment enables EWSR1-FLI1 retargeting.

We first tested for the enrichment of repeat classes in accessible chromatin in tumor cells and primary BM-MSC and found that they shared a high degree of enrichment at simple repeats, relative to other repetitive element classes (Figure 6A). Since EWSR1-FLI1 selectively retargets GGAA-containing simple repeats, we then examined FAIRE signal in BM-MSC and Ewing sarcoma cells at all simple repeats containing this motif, clustering these regions based on their signal in the cancer cells (Figure 6B). We observed a striking similarity in the pattern of chromatin accessibility between the stem and cancer cells. In BM-MSC, the signal was center-weighted at about half of the regions (Figures 6B and S6C). For others, regions flanking the repeat demonstrated the greatest signal.

To explore the connection between chromatin accessibility and EWSR1-FLI1 targeting, we compared FAIRE signal in BM-MSC with EWSR1-FLI1 ChIP signal from Ewing sarcoma cells. Repeats with the greatest FAIRE signal in BM-MSC demonstrated the greatest ChIP signal in the tumor cells (Figure S6D). Similarly, EWSR1-FLI1 targeted those regions for which the maximal FAIRE signal was over the repeat in BM-MSC ($p < 0.001$, permutation). We then explored the activity of EWSR1-FLI1 on chromatin. We compared the difference in FAIRE signal between BM-MSC and the tumor cells with EWSR1-FLI1 ChIP signal. We found a significant correlation between oncoprotein binding and changes in FAIRE signal ($r = 0.74$) (Figure 6C). Taken together, these data lend chromatin-based evidence of an MSC origin for these tumors and, further, demonstrate that characteristics of chromatin MSC predict EWSR1-FLI1 oncoprotein targeting in tumor cells.

We then explored chromatin accessibility using enzymatic approaches. Neither DNase-seq data that we generated from BM-MSC nor published DNase and ATAC data from these cells identified signal enrichment at regions ultimately targeted by EWSR1-FLI1 (Figure 6D) (Patel et al., 2012; Riggi et al., 2014). The absence of signal is consistent with our result in hESC (Figure 3A). In contrast, in Ewing Sarcoma cells these regions were detectable by DNase and ATAC. Moreover, in BM-MSC, ATAC enrichment was noted at these sites only after EWSR1-FLI1 was transduced (Figure 6D and (Riggi et al., 2014). Neither DNase nor ATAC signal enrichment was observed at similar repeats that did not bind EWSR1-FLI1. These data suggest that EWSR1-FLI1 targets nucleosome-destabilized regions which ultimately leads to nucleosome eviction.

Since EWSR1-FLI1 targets a subset of GGAA-containing simple repeats, we asked whether there were other chromatin features that correlated with increased FAIRE signal in BM-MSC and the ability to bind EWSR1-FLI1. Using ChIP from H1-MSC (Bernstein et al., 2010) we examined histone modifications at those sites that are targeted by EWSR1-FLI1 in cancer cells. Of histone modifications available for analysis, we noted a subtle increase in enrichment for H3K14ac, H4K91ac, H2BK12ac, all marks enriched in simple repeats in hESC (Figure 6E). These data suggest that chromatin modifications at critical sites specific to stem cells facilitate EWSR1-FLI1 targeting.

Discussion

By integrating complementary genome-wide approaches we identified a unique chromatin environment in stem cells marked by accessible chromatin at repetitive DNA sequences. Further, we associate these features with the selective targeting of the central oncogene in Ewing sarcoma suggesting that stem cells harbor a permissive environment that facilitates the critical oncogenic step in this cancer.

Though FAIRE-seq revealed the magnitude of this unexpected chromatin signature in hESC, complementary experimental approaches supported this observation. The importance of repetitive elements in stem cells had been described in several recent publications. (Fort et al., 2014; Goke et al., 2015; Grow et al., 2015; Lu et al., 2014; Wang et al., 2014). Transcription of endogenous retroviruses, notably HERVK and HERVH, plays a role in the maintenance of pluripotency. An examination of chromatin in murine ESC using FAIRE

similarly demonstrated variation in regions associated with developmental pathways (Murtha et al., 2015). Although not addressed in this study, our analysis of these data also demonstrated enrichment of repetitive elements in mESC compared to MEF (data not shown). That hESCs had significantly more FAIRE peaks and a generally lower signal-to-noise ratio than differentiated cells suggests decreased chromatin condensation, consistent with biochemical and microscopic approaches (Meshorer et al., 2006; Ricci et al., 2015).

The most characteristic feature associated with accessible repetitive elements was histone acetylation. Variations in histone acetylation have been linked to stem cell differentiation, and nucleosome acetylation can destabilize DNA-nucleosome interactions (Lee et al., 2004; Shogren-Knaak et al., 2006). Interestingly, the sites of acetylation enriched at simple repeats differ from the well-studied H3K27ac and H3K9ac. Segmentation analysis of stem cells has generally categorized repeats as heterochromatic, however these modeling approaches have not included atypical marks, such as H2AK5ac. Indeed, evidence suggests that H2AK5ac enrichment is associated with active regions of chromatin (Cuddapah et al., 2009). Given the paucity of available datasets, features other than histone acetylation may also influence chromatin accessibility. As a functional readout of chromatin states, the inclusion of FAIRE may increase the power of predictive genomic segmentation.

In support of a functional role for repetitive elements, variation in chromatin organization that accompanied differentiation was significantly linked to the regulation of relevant genes. Genes associated with SINEs and simple repeats that demonstrated differential accessibility exhibited pathway enrichment specific to pluripotency such as SOX2, OCT4, and Nanog targets. Similarly, pathways associated with mesenchymal differentiation and function were enriched among regions with gains in accessibility during differentiation. Interestingly, variation in histone posttranslational modifications between induced pluripotent stem cells (iPSC) and ESCs has been inconsistently identified (Guenther et al., 2010; Hawkins et al., 2010). Analysis of iPSC by FAIRE would identify whether features of chromatin accessibility at repetitive elements are restored during reprogramming and could contribute to chromatin-based exploration of the reprogramming process.

The observation of phased nucleosomes flanking all repetitive elements was also unexpected. Ordered nucleosomes were observed even in the absence of FAIRE enrichment and in differentiated cells. Stretches of specific repeated sequences can bend DNA which may attract nucleosomes, and DNA sequence content can influence nucleosome position (Brukner et al., 1993; Hsieh and Griffith, 1988; Lowary and Widom, 1998; Valouev et al., 2011). Our results greatly extend previous reports suggesting that Alu repeats may serve to pattern nucleosomes (Englander and Howard, 1995; Tanaka et al., 2010).

A striking result of our study was the extensive similarity between MSC chromatin and that of Ewing sarcoma. The shared chromatin pattern strongly supports tumor development from a stem-like population, an observation consistent with studies describing similarities in gene expression and capacity for *in vitro* differentiation (Suva et al., 2009; Tirode et al., 2007). Further, our results offer a mechanistic explanation for the cell-type specific targeting of EWSR1-FLI1 in tumor cells. The absence of accessible chromatin at repetitive elements in differentiated cell types may restrict EWSR1-FLI1 targeting offering an explanation of why

this oncogene fails to broadly transform cells (Owen and Lessnick, 2006). Simple repeats, when bound by EWSR1-FLI1, gain enhancer activity to regulate the transcription of multiple genes known to be important for Ewing Sarcoma (Gangwal et al., 2008; Grunewald et al., 2015; Kinsey et al., 2006; Luo et al., 2009; Patel et al., 2012; Smith et al., 2006). Further, germline variation in repetitive element composition has recently been associated with disease risk (Beck et al., 2012; Grunewald et al., 2015). The differences in location and composition of these repetitive regions relative to critical genes across species may partially explain the challenge in generating an animal model that faithfully recapitulates features of Ewing sarcoma (Lin et al., 2008).

Finally, our study offered unexpected technical insights into FAIRE. Previous studies have noted discrepancies between FAIRE and DNase, particularly at distal regulatory elements (Song et al., 2011). However, the biochemical differences characterizing those regions that are enriched by FAIRE but not detected by DNase have not been identified. Compared with DNase and ATAC, FAIRE-seq seems unique in its ability to identify unstable nucleosome-bound regions. Resulting from chromatin organizational differences or histone acetylation, these destabilized nucleosomes may not survive the biochemical extraction process of FAIRE. In a similar fashion, unstable H2A.Z/H3.3 containing nucleosomes have been found at regions deemed “nucleosome depleted” (Jin et al., 2009). In contrast, DNase and ATAC depend on exposed DNA for enzymatic cleavage. Consistent with this difference, DNase and ATAC data from Ewing Sarcoma indicates nucleosome eviction. Nucleosome eviction was also reflected by quantitative gains in FAIRE signal. Strategies that explore chromatin organization yield distinct insights. Apparent differences between these methods may indicate specific states that influence chromatin accessibility.

Overall, we identify a link between stem cell-specific chromatin features at repetitive elements and cancer development. Because of their abundance these elements may broadly influence nucleosome positioning and chromatin remodeling during differentiation. Multiple mechanisms result in variation in repeat element structure and location. How these factors converge to alter chromatin organization will continue to enhance our understanding of development and disease.

Materials and Methods

Cell Culture

The human embryonic stem cell line H1-ESC, H7-ESC, H9-ESC were obtained from WiCell Research Institute (Madison, WI). H1-ESC hES cells were maintained undifferentiated on 6-well plates coated with growth factor-reduced Matrigel (BD Biosciences) in mTeSR1 media (StemCell Technologies) and the media was changed daily. Cells were passaged every three days. To differentiation embryonic stem cells into MSC H1-ESC cells were expanded to 10 cm Matrigel-coated tissue culture dishes. At approximately 80% confluence, cells were dissociated and grown as embryoid bodies which were dissociated and cultured in Mesencult Proliferation Kit (StemCell Technologies).

Primary MSC were isolated from patient bone marrow (UNC IRB exemption 09-0127) using a Ficoll gradient. The mononuclear fraction was isolated and plated. After 3-4 days

adherent cells were selected by magnetic depletion according to manufacturer's instructions (MACS) with CD11b/Mac-1 (BD #555387) and CD45 (BD #555481). The negative fraction consisting of MSCs was then plated and expanded for analysis. HKC were passaged twice a week or media changed every 2 days if not confluent.

FAIRE-Seq

Chromatin from embryonic stem cells was prepared as previously described (Simon et al., 2012). Sequencing libraries were generated from DNA enriched by FAIRE and sequenced (Illumina). 50-bp single end reads were filtered using TagDust (Lassmann et al., 2009) and aligned to hg19 using Bowtie (Langmead et al., 2009). Peaks were called using MACS2 (Zhang et al., 2008). Data are available from GEO (Accession number GSE75172). FASTQ files from HUVEC, NHEK, and K562 FAIRE were obtained from ENCODE and processed as above.

Genomic Window Analysis, segmentation and peak detection

Z-scores were calculated for each sample. Scores were then averaged over 500 bp non-overlapping windows. Significance was assayed between hESCs and HUVEC, NHEK, and K562 by t-test (p-value < 0.01,) and clustered. Data was clustered and visualized. Significantly altered windows were then intersected with ChromHMM segmentation coordinates obtained from ENCODE. Peak summits, as defined by MACS2, were extended ± 1 bp and intersected with known repetitive elements from the Repeat Masker track (UCSC).

ChIP-Seq data and analysis

EWSR1-FLI1 data was obtained from previously published study (Patel et al., 2012). Alignment data for H1-ESC and H1-MSC histone modifications were downloaded from the Epigenome Roadmap and processed into single-bp wiggle files using ZINBA (Rashid et al., 2011). Signal from histone modifications at simple repeats and SINEs that intersected H1-ESC FAIRE peaks (FAIRE-enriched) or an equal number of FAIRE-negative repeats were compared by summing the ChIP signal ± 250 bp from center of the simple repeat or the length of the SINEs. Enrichment ranking was determined by ordering on the difference between FAIRE-enriched and FAIRE-negative regions. Ranking for control regions TSS and CTCF was determined by summing signal +− 500 bp from center of respective feature.

Salt Fractionation

H1-ESCs and HKCs were cultured in normal growth conditions. 10 million cells were counted (Bio-Rad TC20) and treated as previously described (Teves and Henikoff, 2012) with the following modifications. 5 U of MNase I was used to digest nuclei. Extractions were performed at 40 mM, 80 mM, 160 mM, 320 mM, and 640 mM NaCl in a volume of 50 μL. For sequencing, H1-ESC and HKC nuclei were sequentially extracted with low NaCl (160 mM) and high NaCl (640 mM) concentrations. DNA was isolated from the fractions as previously described (Teves and Henikoff, 2012) and prepared for sequencing (Illumina). Paired-end reads were mapped to hg19 to map nucleosome positioning.

Immunoblot of Salt Fractions

Proteins eluted during salt fractionation were aspirated onto nitrocellulose membranes according the manufacturer's instructions (Bio-Dot SF, Bio-Rad). The membranes were immunoblotted using antibodies for H2AK5ac (Abcam #1764) and pan-H3 (Courtesy of B. Strahl) and imaged using an infrared secondary (Li-Cor).

DNase Hypersensitivity I

DNase data for H1-ESC was downloaded from the ENCODE. DNase I hypersensitivity was performed on primary BM-MSC as previously described (Crawford et al., 2006) and sequenced (Illumina). Resulting sequencing reads were in-silico clipped to 20 bp and aligned to the hg19.

RNA-Sequencing and analysis

RNA was isolated from BM-MSCs using Trizol according to manufacturer's directions. Ribosomal RNA depletion and preparation for sequencing were performed as described previously (Simon et al., 2014). Sequencing reads were aligned to hg19 using TopHat (Trapnell et al., 2010). RPKM were calculated and used for comparison against H1-ESC. H1-ESC RNA-seq data was downloaded from the ENCODE Consortium.

Chromatin remodeling during differentiation

Average Z-score of FAIRE signal was computed for each SINEs and simple repeat in hESCs, BM-MSCs, H1-MSCs, and K562, NHEK, and HUVEC. Regions significantly different (p-value < 0.01 for SINEs and p-value < 0.05 for simple repeat, t-test) between hESCs and BM-MSCs were filtered so that maximum-minimum value > 1. Regions passing this threshold were then biclustered including data from H1-MSC, NHEK, HUVEC, and K562. Repeats in identified clusters were then associated with the closest gene transcription start site within 50 kb. Only genes classified as differentially regulated (ESC/MSC RPKM >2 fold-change) were considered. Fraction of genes that were either up in ESC or up in MSC were then counted. Gene Ontology enrichment was performed with gene lists from each chromatin category using Piano (Varemo et al., 2013). Top terms from each significant cluster (adjusted p-value < 0.05) were selected for presentation; all results are reported in Table S4. For permutations, p-values associated with repeats were scrambled and re-selected to $p < 0.01$ for SINEs and $p < 0.05$ for Simple Repeats 1,000 times. The proportion of up- and down-regulated genes associated with those regions was computed and averaged. Error bars represent standard deviation across all 1,000 permutations.

Statistics

Statistical analyses were performed using student's t-test and/or random permutation. Relevant p-values are indicated in each panel.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgement

We gratefully acknowledge the assistance with DNase from A. Safi and G. Crawford. We thank B. Strahl for antibodies and advice. We also acknowledge J. Lieb and members of the Lieb lab, A. Spagnoli and members of the Spagnoli lab, and T. Furey for advice. We acknowledge the assistance of the UNC Lineberger High Throughput Sequencing (P. Mieczkowski), Flow Cytometry (N. Fisher) and Pluripotent Stem Cell core. We gratefully acknowledge support from the National Institutes of Health R01CA166447 to IJD, P30CA016086 to UNC Lineberger, the University Cancer Research Fund, V Foundation for Cancer Research, the Wide Open Foundation and the Corn-Hammond Fund for Pediatric Oncology. NG was supported by T32GM007092 and Initiative for Maximizing Student Diversity (IMSD) 5R25GM055336.

References

Beck R, Monument MJ, Watkins WS, Smith R, Boucher KM, Schiffman JD, Jorde LB, Randall RL, Lessnick SL. EWS/FLI-responsive GGAA microsatellites exhibit polymorphic differences between European and African populations. Cancer genetics. 2012; 205:304–312. [PubMed: 22749036]

Bernstein BE, Mikkelsen TS, Xie X, Kamal M, Huebert DJ, Cuff J, Fry B, Meissner A, Wernig M, Plath K, et al. A bivalent chromatin structure marks key developmental genes in embryonic stem cells. Cell. 2006; 125:315–326. [PubMed: 16630819]

Bernstein BE, Stamatoyannopoulos JA, Costello JF, Ren B, Milosavljevic A, Meissner A, Kellis M, Marra MA, Beaudet AL, Ecker JR, et al. The NIH Roadmap Epigenomics Mapping Consortium. Nature biotechnology. 2010; 28:1045–1048.

Boyer LA, Lee TI, Cole MF, Johnstone SE, Levine SS, Zucker JP, Guenther MG, Kumar RM, Murray HL, Jenner RG, et al. Core transcriptional regulatory circuitry in human embryonic stem cells. Cell. 2005; 122:947–956. [PubMed: 16153702]

Boyle AP, Davis S, Shulha HP, Meltzer P, Margulies EH, Weng Z, Furey TS, Crawford GE. High-resolution mapping and characterization of open chromatin across the genome. Cell. 2008; 132:311–322. [PubMed: 18243105]

Brukner I, Dlakic M, Savic A, Susic S, Pongor S, Suck D. Evidence for opposite groove-directed curvature of GGGCCC and AAAAA sequence elements. Nucleic acids research. 1993; 21:1025–1029. [PubMed: 8451169]

Buenrostro JD, Giresi PG, Zaba LC, Chang HY, Greenleaf WJ. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. Nature methods. 2013; 10:1213–1218. [PubMed: 24097267]

Consortium EP, Dunham I, Kundaje A, Aldred SF, Collins PJ, Davis CA, Doyle F, Epstein CB, Frietze S, Harrow J, et al. An integrated encyclopedia of DNA elements in the human genome. Nature. 2012; 489:57–74. [PubMed: 22955616]

Crawford GE, Holt IE, Whittle J, Webb BD, Tai D, Davis S, Margulies EH, Chen Y, Bernat JA, Ginsburg D, et al. Genome-wide mapping of DNase hypersensitive sites using massively parallel signature sequencing (MPSS). Genome research. 2006; 16:123–131. [PubMed: 16344561]

Cuddapah S, Jothi R, Schones DE, Roh TY, Cui K, Zhao K. Global analysis of the insulator binding protein CTCF in chromatin barrier regions reveals demarcation of active and repressive domains. Genome research. 2009; 19:24–32. [PubMed: 19056695]

Dixon JR, Jung I, Selvaraj S, Shen Y, Antosiewicz-Bourget JE, Lee AY, Ye Z, Kim A, Rajagopal N, Xie W, et al. Chromatin architecture reorganization during stem cell differentiation. Nature. 2015; 518:331–336. [PubMed: 25693564]

Efroni S, Duttagupta R, Cheng J, Dehghani H, Hoeppner DJ, Dash C, Bazett-Jones DP, Le Grice S, McKay RD, Buetow KH, et al. Global transcription in pluripotent embryonic stem cells. Cell stem cell. 2008; 2:437–447. [PubMed: 18462694]

Englander EW, Howard BH. Nucleosome positioning by human Alu elements in chromatin. The Journal of biological chemistry. 1995; 270:10091–10096. [PubMed: 7730313]

Ernst J, Kellis M. Discovery and characterization of chromatin states for systematic annotation of the human genome. Nature biotechnology. 2010; 28:817–825.

Ernst J, Kellis M. ChromHMM: automating chromatin-state discovery and characterization. Nature methods. 2012; 9:215–216. [PubMed: 22373907]

Fort A, Hashimoto K, Yamada D, Salimullah M, Keya CA, Saxena A, Bonetti A, Voineagu I, Bertin N, Kratz A, et al. Deep transcriptome profiling of mammalian stem cells supports a regulatory role for retrotransposons in pluripotency maintenance. Nature genetics. 2014; 46:558–566. [PubMed: 24777452]

Fu Y, Sinha M, Peterson CL, Weng Z. The insulator binding protein CTCF positions 20 nucleosomes around its binding sites across the human genome. PLoS genetics. 2008; 4:e1000138. [PubMed: 18654629]

Gangwal K, Sankar S, Hollenhorst PC, Kinsey M, Haroldsen SC, Shah AA, Boucher KM, Watkins WS, Jorde LB, Graves BJ, et al. Microsatellites as EWS/FLI response elements in Ewing's sarcoma. Proceedings of the National Academy of Sciences of the United States of America. 2008; 105:10149–10154. [PubMed: 18626011]

Giresi PG, Kim J, McDaniell RM, Iyer VR, Lieb JD. FAIRE (Formaldehyde-Assisted Isolation of Regulatory Elements) isolates active regulatory elements from human chromatin. Genome research. 2007; 17:877–885. [PubMed: 17179217]

Goke J, Lu X, Chan YS, Ng HH, Ly LH, Sachs F, Szczerbinska I. Dynamic transcription of distinct classes of endogenous retroviral elements marks specific populations of early human embryonic cells. Cell stem cell. 2015; 16:135–141. [PubMed: 25658370]

Grow EJ, Flynn RA, Chavez SL, Bayless NL, Wossidlo M, Wesche DJ, Martin L, Ware CB, Blish CA, Chang HY, et al. Intrinsic retroviral reactivation in human preimplantation embryos and pluripotent cells. Nature. 2015; 522:221–225. [PubMed: 25896322]

Grunewald TG, Bernard V, Gilardi-Hebenstreit P, Raynal V, Surdez D, Aynaud MM, Mirabeau O, Cidre-Aranaz F, Tirode F, Zaidi S, et al. Chimeric EWSR1-FLI1 regulates the Ewing sarcoma susceptibility gene EGR2 via a GGAA microsatellite. Nature genetics. 2015; 47:1073–1078. [PubMed: 26214589]

Guenther MG, Frampton GM, Soldner F, Hockemeyer D, Mitalipova M, Jaenisch R, Young RA. Chromatin structure and gene expression programs of human embryonic and induced pluripotent stem cells. Cell stem cell. 2010; 7:249–257. [PubMed: 20682450]

Hawkins RD, Hon GC, Lee LK, Ngo Q, Lister R, Pelizzola M, Edsall LE, Kuan S, Luu Y, Klugman S, et al. Distinct epigenomic landscapes of pluripotent and lineage-committed human cells. Cell stem cell. 2010; 6:479–491. [PubMed: 20452322]

Henikoff S, Henikoff JG, Sakai A, Loeb GB, Ahmad K. Genome-wide profiling of salt fractions maps physical properties of chromatin. Genome research. 2009; 19:460–469. [PubMed: 19088306]

Hsieh CH, Griffith JD. The terminus of SV40 DNA replication and transcription contains a sharp sequence-directed curve. Cell. 1988; 52:535–544. [PubMed: 2830026]

Jin C, Zang C, Wei G, Cui K, Peng W, Zhao K, Felsenfeld G. H3.3/H2A.Z double variant-containing nucleosomes mark ‘nucleosome-free regions’ of active promoters and other regulatory regions. Nature genetics. 2009; 41:941–945. [PubMed: 19633671]

Kafer GR, Lehnert SA, Pantaleon M, Kaye PL, Moser RJ. Expression of genes coding for histone variants and histone-associated proteins in pluripotent stem cells and mouse preimplantation embryos. Gene expression patterns : GEP. 2010; 10:299–305. [PubMed: 20601166]

Kinsey M, Smith R, Lessnick SL. NR0B1 is required for the oncogenic phenotype mediated by EWS/FLI in Ewing's sarcoma. Molecular cancer research : MCR. 2006; 4:851–859. [PubMed: 17114343]

Kurimoto K, Yabuta Y, Hayashi K, Ohta H, Kiyonari H, Mitani T, Moritoki Y, Kohri K, Kimura H, Yamamoto T, et al. Quantitative Dynamics of Chromatin Remodeling during Germ Cell Specification from Mouse Embryonic Stem Cells. Cell stem cell. 2015; 16:517–532. [PubMed: 25800778]

Langmead B, Trapnell C, Pop M, Salzberg SL. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome biology. 2009; 10:R25. [PubMed: 19261174]

Lassmann T, Hayashizaki Y, Daub CO. TagDust--a program to eliminate artifacts from next generation sequencing data. Bioinformatics. 2009; 25:2839–2840. [PubMed: 19737799]

Lee JH, Hart SR, Skalnik DG. Histone deacetylase activity is required for embryonic stem cell differentiation. Genesis (New York, NY : 2000). 2004; 38:32–38.

Lin PP, Pandey MK, Jin F, Xiong S, Deavers M, Parant JM, Lozano G. EWS-FLI1 induces developmental abnormalities and accelerates sarcoma formation in a transgenic mouse model. Cancer research. 2008; 68:8968–8975. [PubMed: 18974141]

Loh YH, Wu Q, Chew JL, Vega VB, Zhang W, Chen X, Bourque G, George J, Leong B, Liu J, et al. The Oct4 and Nanog transcription network regulates pluripotency in mouse embryonic stem cells. Nature genetics. 2006; 38:431–440. [PubMed: 16518401]

Lowary PT, Widom J. New DNA sequence rules for high affinity binding to histone octamer and sequence-directed nucleosome positioning. Journal of molecular biology. 1998; 276:19–42. [PubMed: 9514715]

Lu X, Sachs F, Ramsay L, Jacques PE, Goke J, Bourque G, Ng HH. The retrovirus HERVH is a long noncoding RNA required for human embryonic stem cell identity. Nature structural & molecular biology. 2014; 21:423–425.

Luo W, Gangwal K, Sankar S, Boucher KM, Thomas D, Lessnick SL. GSTM4 is a microsatellite-containing EWS/FLI target involved in Ewing's sarcoma oncogenesis and therapeutic resistance. Oncogene. 2009; 28:4126–4132. [PubMed: 19718047]

Meshorer E, Yellajoshula D, George E, Scambler PJ, Brown DT, Misteli T. Hyperdynamic plasticity of chromatin proteins in pluripotent embryonic stem cells. Developmental cell. 2006; 10:105–116. [PubMed: 16399082]

Mikkelsen TS, Ku M, Jaffe DB, Issac B, Lieberman E, Giannoukos G, Alvarez P, Brockman W, Kim TK, Koche RP, et al. Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. Nature. 2007; 448:553–560. [PubMed: 17603471]

Murtha M, Strino F, Tokcaer-Keskin Z, Sumru Bayin N, Shalabi D, Xi X, Kluger Y, Dailey L. Comparative FAIRE-seq Analysis Reveals Distinguishing Features of the Chromatin Structure of Ground State- and Primed-Pluripotent Cells. Stem cells (Dayton, Ohio). 2015; 33:378–391.

Orvis T, Hepperla A, Walter V, Song S, Simon J, Parker J, Wilkerson MD, Desai N, Major MB, Hayes DN, et al. BRG1/SMARCA4 inactivation promotes non-small cell lung cancer aggressiveness by altering chromatin organization. Cancer research. 2014; 74:6486–6498. [PubMed: 25115300]

Owen LA, Lessnick SL. Identification of target genes in their native cellular context: an analysis of EWS/FLI in Ewing's sarcoma. Cell cycle. 2006; 5:2049–2053. [PubMed: 16969112]

Paige SL, Thomas S, Stoick-Cooper CL, Wang H, Maves L, Sandstrom R, Pabon L, Reinecke H, Pratt G, Keller G, et al. A temporal chromatin signature in human embryonic stem cells identifies regulators of cardiac development. Cell. 2012; 151:221–232. [PubMed: 22981225]

Patel M, Simon JM, Iglesia MD, Wu SB, McFadden AW, Lieb JD, Davis IJ. Tumor-specific retargeting of an oncogenic transcription factor chimera results in dysregulation of chromatin and transcription. Genome research. 2012; 22:259–270. [PubMed: 22086061]

Rashid NU, Giresi PG, Ibrahim JG, Sun W, Lieb JD. ZINBA integrates local covariates with DNA-seq data to identify broad and narrow regions of enrichment, even within amplified genomic regions. Genome biology. 2011; 12:R67. [PubMed: 21787385]

Ricci MA, Manzo C, Garcia-Parajo MF, Lakadamyali M, Cosma MP. Chromatin fibers are formed by heterogeneous groups of nucleosomes in vivo. Cell. 2015; 160:1145–1158. [PubMed: 25768910]

Riggi N, Knoechel B, Gillespie SM, Rheinbay E, Boulay G, Suva ML, Rossetti NE, Boonseng WE, Oksuz O, Cook EB, et al. EWS-FLI1 Utilizes Divergent Chromatin Remodeling Mechanisms to Directly Activate or Repress Enhancer Elements in Ewing Sarcoma. Cancer cell. 2014; 26:668–681. [PubMed: 25453903]

Rodriguez R, Rubio R, Menendez P. Modeling sarcomagenesis using multipotent mesenchymal stem cells. Cell research. 2012; 22:62–77. [PubMed: 21931359]

Sanders MM. Fractionation of nucleosomes by salt elution from micrococcal nuclease-digested nuclei. The Journal of cell biology. 1978; 79:97–109. [PubMed: 701381]

Shogren-Knaak M, Ishii H, Sun JM, Pazin MJ, Davie JR, Peterson CL. Histone H4-K16 acetylation controls chromatin structure and protein interactions. Science (New York, NY). 2006; 311:844–847.

Simon JM, Giresi PG, Davis IJ, Lieb JD. Using formaldehyde-assisted isolation of regulatory elements (FAIRE) to isolate active regulatory DNA. Nature protocols. 2012; 7:256–267. [PubMed: 22262007]

Simon JM, Hacker KE, Singh D, Brannon AR, Parker JS, Weiser M, Ho TH, Kuan PF, Jonasch E, Furey TS, et al. Variation in chromatin accessibility in human kidney cancer links H3K36 methyltransferase loss with widespread RNA processing defects. Genome research. 2014; 24:241–250. [PubMed: 24158655]

Smit AFA, Hubley R, Green P. RepeatMasker Open-4.0. 2015

Smith R, Owen LA, Trem DJ, Wong JS, Whangbo JS, Golub TR, Lessnick SL. Expression profiling of EWS/FLI identifies NKX2.2 as a critical target gene in Ewing's sarcoma. Cancer cell. 2006; 9:405–416. [PubMed: 16697960]

Song L, Zhang Z, Grasfeder LL, Boyle AP, Giresi PG, Lee BK, Sheffield NC, Graf S, Huss M, Keefe D, et al. Open chromatin defined by DNaseI and FAIRE identifies regulatory elements that shape cell-type identity. Genome research. 2011; 21:1757–1767. [PubMed: 21750106]

Suva ML, Riggi N, Stehle JC, Baumer K, Tercier S, Joseph JM, Suva D, Clement V, Provero P, Cironi L, et al. Identification of cancer stem cells in Ewing's sarcoma. Cancer research. 2009; 69:1776–1781. [PubMed: 19208848]

Tanaka Y, Yamashita R, Suzuki Y, Nakai K. Effects of Alu elements on global nucleosome positioning in the human genome. BMC genomics. 2010; 11:309. [PubMed: 20478020]

Teves SS, Henikoff S. Salt fractionation of nucleosomes for genome-wide profiling. Methods in molecular biology. 2012; 833:421–432. [PubMed: 22183608]

Thomson JA, Itskovitz-Eldor J, Shapiro SS, Waknitz MA, Swiergiel JJ, Marshall VS, Jones JM. Embryonic stem cell lines derived from human blastocysts. Science (New York, NY). 1998; 282:1145–1147.

Tirode F, Laud-Duval K, Prieur A, Delorme B, Charbord P, Delattre O. Mesenchymal stem cell features of Ewing tumors. Cancer cell. 2007; 11:421–429. [PubMed: 17482132]

Trapnell C, Williams BA, Pertea G, Mortazavi A, Kwan G, van Baren MJ, Salzberg SL, Wold BJ, Pachter L. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. Nature biotechnology. 2010; 28:511–515.

Valouev A, Johnson SM, Boyd SD, Smith CL, Fire AZ, Sidow A. Determinants of nucleosome organization in primary human cells. Nature. 2011; 474:516–520. [PubMed: 21602827]

Varemo L, Nielsen J, Nookaew I. Enriching the gene set analysis of genome-wide data by incorporating directionality of gene expression and combining statistical hypotheses and methods. Nucleic acids research. 2013; 41:4378–4391. [PubMed: 23444143]

Wamstad JA, Alexander JM, Truty RM, Shrikumar A, Li F, Eilertson KE, Ding H, Wylie JN, Pico AR, Capra JA, et al. Dynamic and coordinated epigenetic regulation of developmental transitions in the cardiac lineage. Cell. 2012; 151:206–220. [PubMed: 22981692]

Wang J, Xie G, Singh M, Ghanbarian AT, Rasko T, Szvetnik A, Cai H, Besser D, Prigione A, Fuchs NV, et al. Primate-specific endogenous retrovirus-driven transcription defines naive-like stem cells. Nature. 2014; 516:405–409. [PubMed: 25317556]

Wen B, Wu H, Shinkai Y, Irizarry RA, Feinberg AP. Large histone H3 lysine 9 dimethylated chromatin blocks distinguish differentiated from embryonic stem cells. Nature genetics. 2009; 41:246–250. [PubMed: 19151716]

West JA, Cook A, Alver BH, Stadtfeld M, Deaton AM, Hochedlinger K, Park PJ, Tolstorukov MY, Kingston RE. Nucleosomal occupancy changes locally over key regulatory regions during cell differentiation and reprogramming. Nature communications. 2014; 5:4719.

Zhang Y, Liu T, Meyer CA, Eeckhoute J, Johnson DS, Bernstein BE, Nusbaum C, Myers RM, Brown M, Li W, et al. Model-based analysis of ChIP-Seq (MACS). Genome biology. 2008; 9:R137. [PubMed: 18798982]

Highlights

- Stem cells are characterized by chromatin accessibility at repetitive elements.
- Accessible repetitive elements are marked by distinct histone modifications.
- Stem cell differentiation induces chromatin remodeling at repetitive elements.
- Stem cell chromatin offers a permissive environment for Ewing sarcoma development.

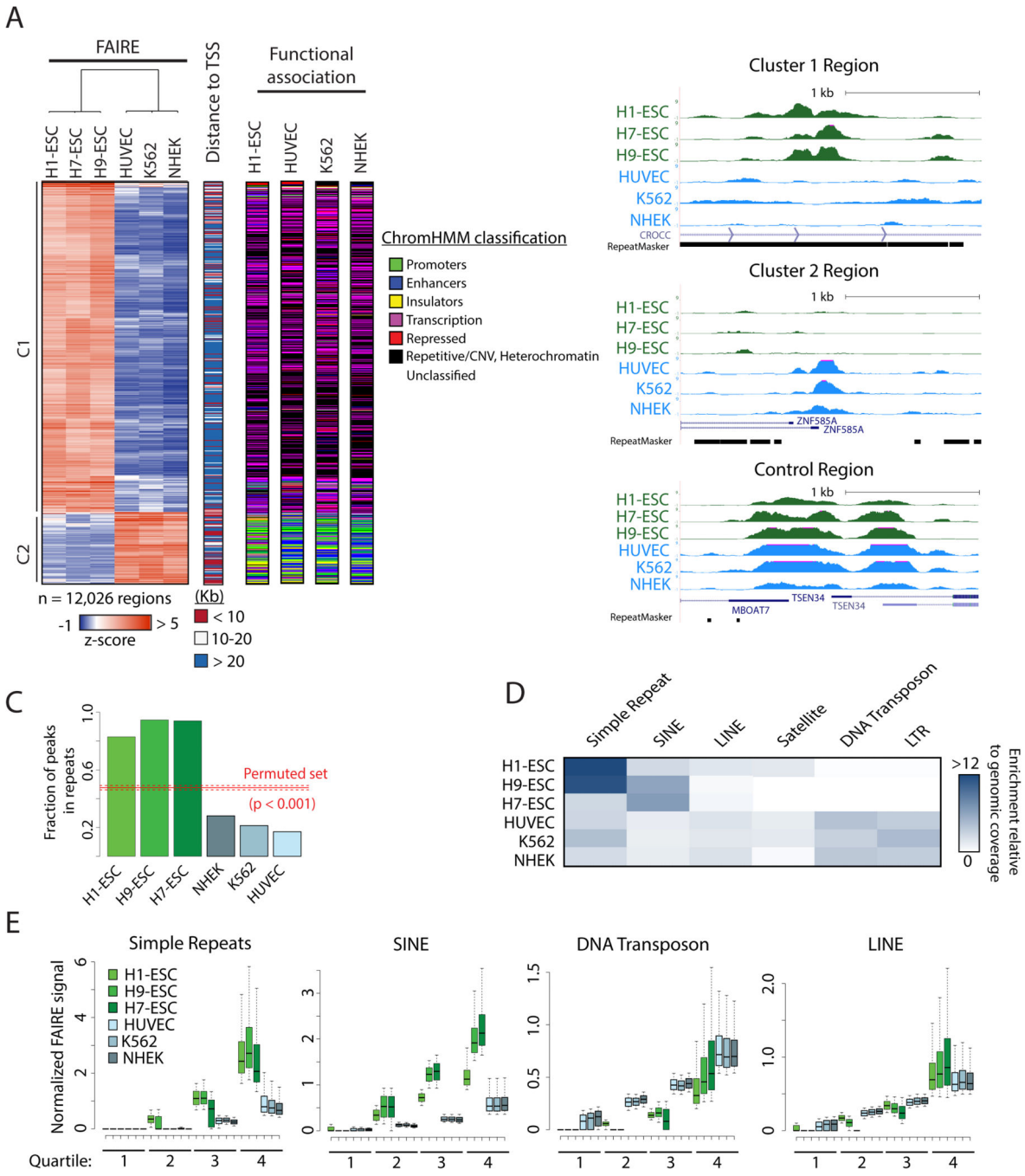


Figure 1. FAIRE-seq enriched regions specific to hESC are dominated by simple repeats and SINEs

(A) Heatmap of those regions with significantly different FAIRE enrichment between hESC and control HUVEC, K562, NHEK (500 bp windows, $p \leq 0.01$, t-test, $row_{max} - row_{min} > 3$). Regions were assigned classes based on distance to nearest TSS (<10 Kb red, 10-20 Kb white, >20 Kb blue and by segmentation analysis (Consortium et al., 2012). See also Supplementary Figure 1.

(B) Genome browser images of representative regions selected from Cluster 1 (top panel), Cluster 2 (middle panel), and a Control region (bottom panel).

(C) Fraction of top 150,000 peak summits that overlapped a repetitive element in hESCs (green) and controls (blue). Fractional overlap with an H1-ESC permuted peak set (red line) and standard deviation (dashed lines) are shown.
(D) Heatmap depicting the enrichment of specific classes of repetitive elements in MACS2-identified FAIRE-enriched regions, relative to genomic coverage.
(E) Normalized FAIRE signal at Simple Repeats, SINEs, DNA Transposon, and LINE in hESC (green) and control cell lines (blue) plotted by quartile.

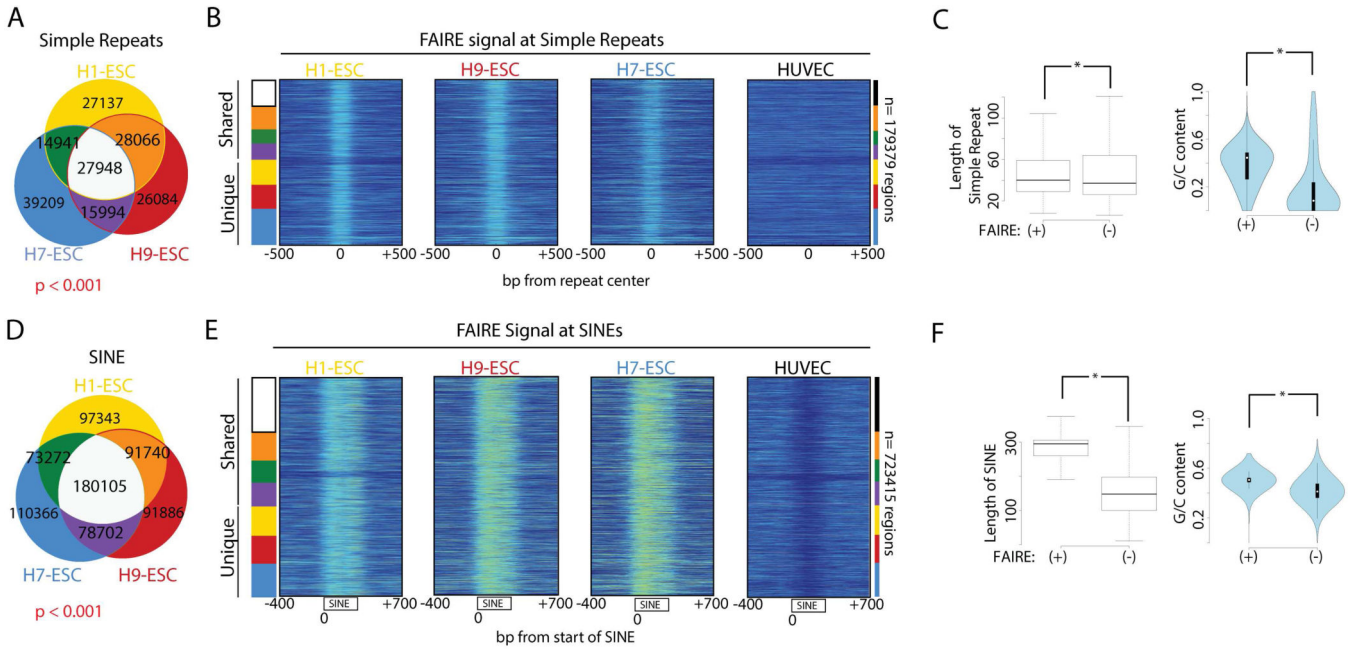


Figure 2. hESC share FAIRE signal enrichment at Simple Repeats and SINEs

(A) The union set of simple repeats with FAIRE signal in the top quartile (Q4) for hESC are shown. ($p < 0.001$, permutation based on all simple repeats, also see Supplementary Figure 3).

(B) Heatmap demonstrating FAIRE signal at simple repeats grouped by categories (colors defined in A) with. Distance represents bp from the center of the simple repeat.

(C) Lengths (left) and G/C content (right) of simple repeats that are FAIRE-enriched (+) in all hESC (see panel A, n = 27,948) and an equal number of randomly selected simple repeats from quartile 1 (FAIRE −).

(D) The union set of SINEs with FAIRE signal in the top quartile (Q4) for hESC are shown. ($p < 0.001$, permutation based on all SINEs, also see Supplementary Figure 3).

(E) Heatmap demonstrating FAIRE signal at SINEs grouped by categories (colors defined in D). Distance represents bp from the start of the SINEs.

(F) Lengths (left) and G/C content (right) of SINEs that are FAIRE-enriched (+) in all hESC (see panel D, n = 180,105) and an equal number of randomly selected simple repeats from quartile 1 (FAIRE −).

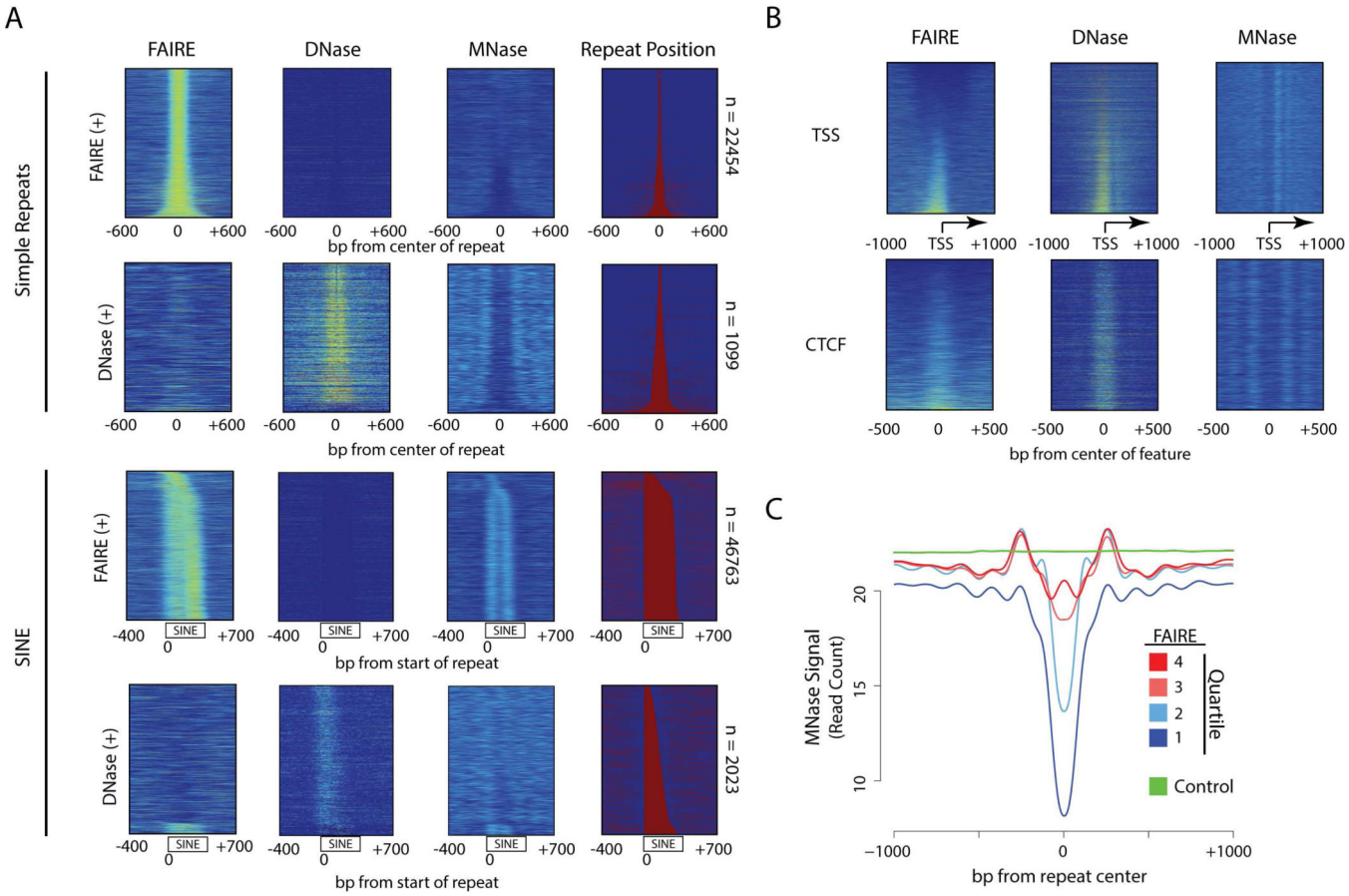


Figure 3. Nucleosome-bound repetitive regions are identified by FAIRE

(A) Heatmap representations of H1-ESC FAIRE-seq, DNase, and MNase-seq signal (West et al., 2014) at FAIRE-enriched (FAIRE +) or DNase-enriched (DNase +) simple repeats and SINEs rank ordered by length. For reference, repeat positions (defined by RepeatMasker) are also plotted.

(B) Heatmaps of H1-ESC FAIRE-seq, DNase, and MNase-seq signal at transcription start sites (TSS) and CTCF sites.

(C) H1-ESC MNase-seq signal at simple repeats grouped by quartiles of FAIRE signal. An equal number of random genomic windows are plotted for comparison (control, green).

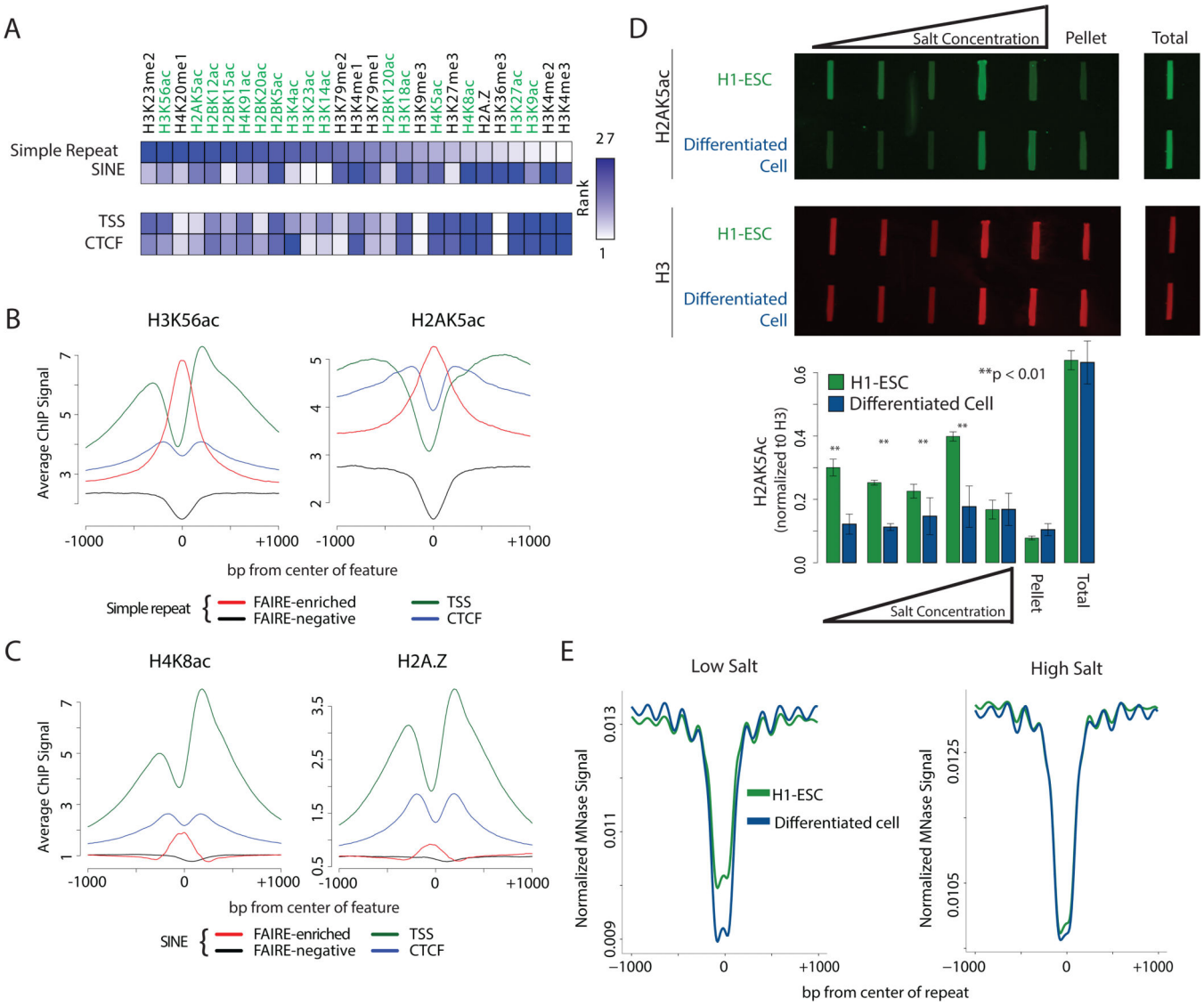


Figure 4. Distinct histone modifications characterize FAIRE-enriched repeats

(A) Heatmap of ranked histone posttranslational modifications. Differential ChIP signal comparing FAIRE-enriched (+) with FAIRE-negative repeats (simple repeat +/− 250 bp from center or SINEs start to +300 bp) was rank ordered. For comparison, signal at TSS or CTCF (+/−500 bp) was rank ordered. Histone modification by acetylation is highlighted (green).

(B) Mean H1-ESC ChIP signal of selected histone posttranslational modifications at H1-ESC FAIRE-enriched (red line) and FAIRE-negative (black line) simple repeats, and control regions (TSS green line, CTCF blue line). Supplemental figure 5 contains all available histone modifications.

(C) Mean H1-ESC ChIP signal of H4K8ac and H2A.Z at H1-ESC FAIRE-enriched (red line) and FAIRE-negative (black line) SINEs, and control regions (TSS green line, CTCF blue line). Supplemental figure 5 contains all available histone modifications.

(D) Salt fractionated nuclear extracts from H1-ESC and HKC (differentiated cell) were immunoblotted with anti-H2AK5ac (green) and anti-pan-H3 (red). Fluorescence intensity was quantified and normalized to H3.

(E) Mean H1-ESC and HKC (differentiated cell) MNase-seq signal from low (left) and high (right) salt fractions at simple repeats. Signal was normalized to reads per million mapped.

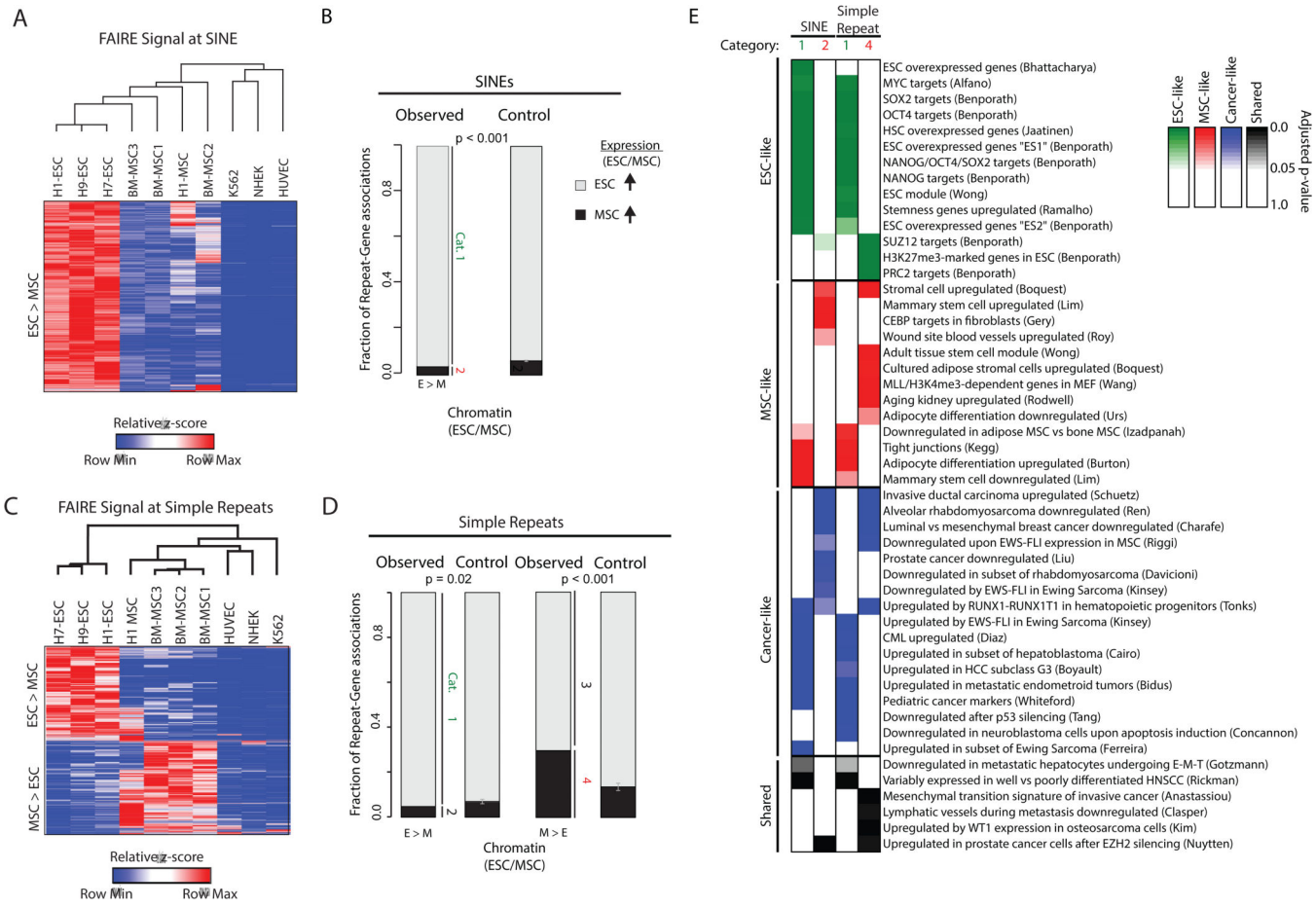


Figure 5. Repetitive elements undergo chromatin remodeling during differentiation

(A) FAIRE signal from hESC, BM-MSC, H1-MSC, K562, NHEK and HUVEC at SINEs characterized by significantly different FAIRE signal between hESCs and BM-MSCs (t-test $p<0.01$, $row_{max} - row_{min} > 1$) were z-score transformed and clustered. Heatmap scale represents relative z-scores.

(B) Fraction of genes linked to variable FAIRE at repeats that demonstrate differential gene expression (increased in hESC – gray, category 1; increased in MSC – black, category 2). Differential expression was defined as RPKM > 2 fold change. Control represents the average from 1000 permutations performed with equal number of selected repeats with permuted p-values. Error bars represent standard deviation of the permutations.

(C) FAIRE signal from hESC, BM-MSC, H1-MSC, K562, NHEK and HUVEC at Simple Repeats characterized by significantly different FAIRE signal between hESCs and BM-MSCs (t-test $p<0.01$, $row_{max} - row_{min} > 1$) were z-score transformed and biclustered. Heatmap scale represents relative z-scores.

(D) Fraction of genes linked to variable FAIRE at repeats that demonstrate differential gene expression. Genes associated with increased FAIRE in hESC are further divided, those associated with genes with increased expression in hESC (gray, category 1) and those associated with genes with increased expression in MSC (black, category 2). Genes associated with increased FAIRE in MSC are also further divided, those associated with genes with increased expression in hESC (gray, category 3) and those associated with genes

with increased expression in MSC (black, category 4). Differential expression was defined as RPKM > 2-fold change. Control represents the average from 1000 permutations performed with equal number of selected repeats with permuted p-values. Error bars represent standard deviation of the permutations.

(E) Gene ontologies enriched for genes identified in category 1 and 2 from (B) and category 1 and 4 from (D). Ontologies were organized by ESC-like (green), MSC-like (red), Cancer-like (blue), and Shared (black). p-value intensity determined by shade of each color. Also see Table S4.

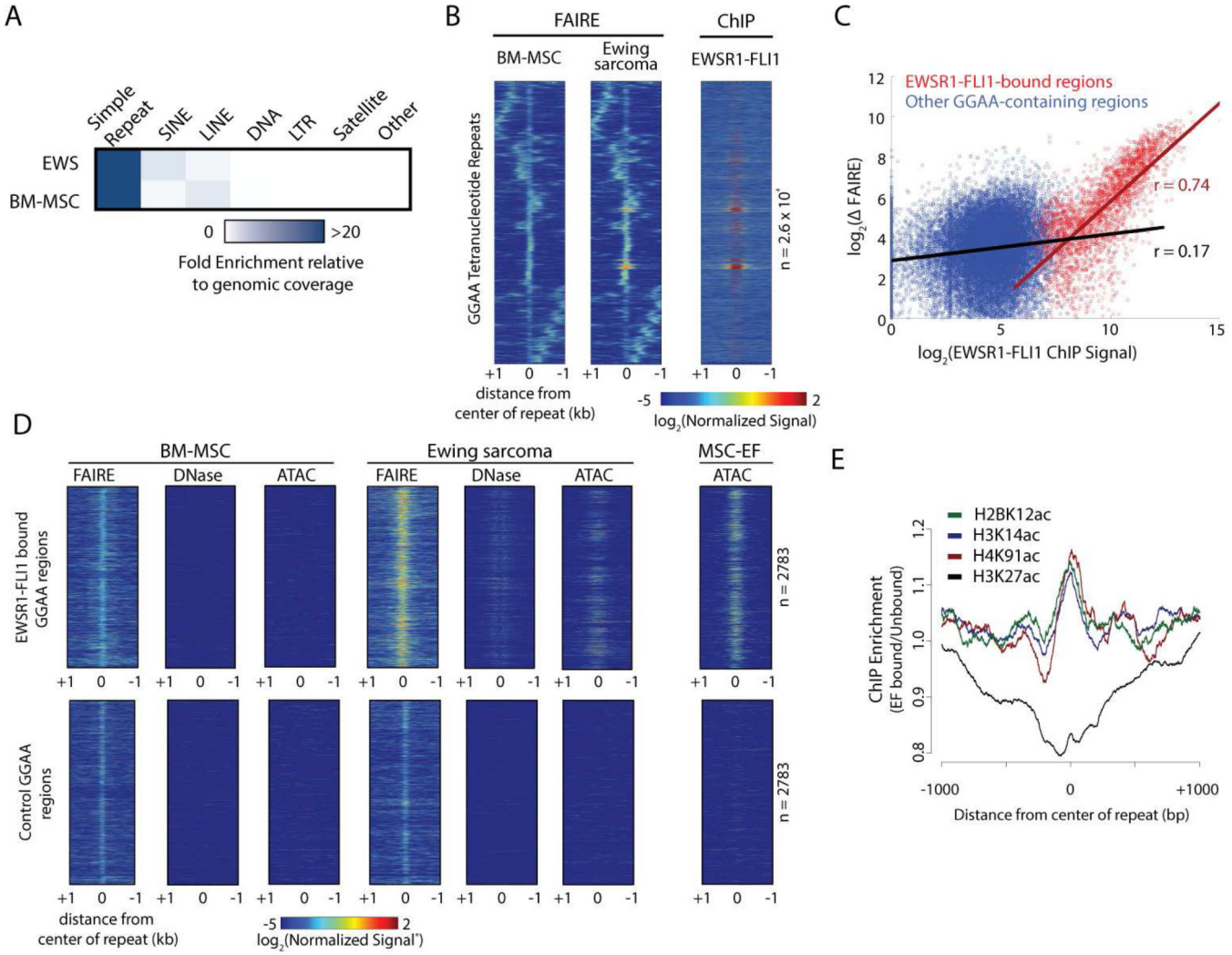


Figure 6. EWSR1-FLI1 oncoprotein exploits the unique chromatin environment of stem cells

(A) Heatmap depicting the enrichment of specific classes of repetitive elements in MACS2-identified FAIRE-enriched regions in Ewing Sarcoma (EWS) and BM-MSC chromatin, relative to genomic coverage.

(B) Clustered BM-MSC or EWS FAIRE signal at all $(GGAA)_n$-containing simple repeats (left). EWSR1-FLI1 ChIP signal in EWS at $(GGAA)_n$-containing simple repeats (right).

(C) Scatterplot of log2 transformed FAIRE change between BM-MSC and EWS and EWSR1-FLI1 ChIP signal at EWSR1-FLI1 bound (red) or unbound (blue) repeats. Pearson correlation shown.

(D) FAIRE, DNase and ATAC signal at EWSR1-FLI1 binding sites in BM-MSC, Ewing Sarcoma (EWS), and MSCs exogenously expressing EWSR1-FLI1 (Riggi et al., 2014). Distance represents Kb from the center of the repeat. FAIRE and DNase data were normalized to overall read count. *ATAC read count was unavailable and consequently not normalized.

(E) Fold change of H1-MSC ChIP signal for H3K14ac, H4K91ac, H2BK12ac, and H3K27ac at repeats bound by EWSR1-FLI1 in Ewing sarcoma cells relative to a equal

number of randomly selected repeats that were not bound. Distance represents from center of repeat.